



به نام خدا



دانشگاه تهران
دانشکده مهندسی برق و کامپیوتر
مقدمه‌ای بر استنباط آماری

گزارش تمرین 2

نام و نام خانوادگی	سید مهدی موسوی
شماره دانشجویی	۸۱۰۱۹۷۵۹۳
تاریخ ارسال گزارش	۱۴۰۲/۱۰/۰۳

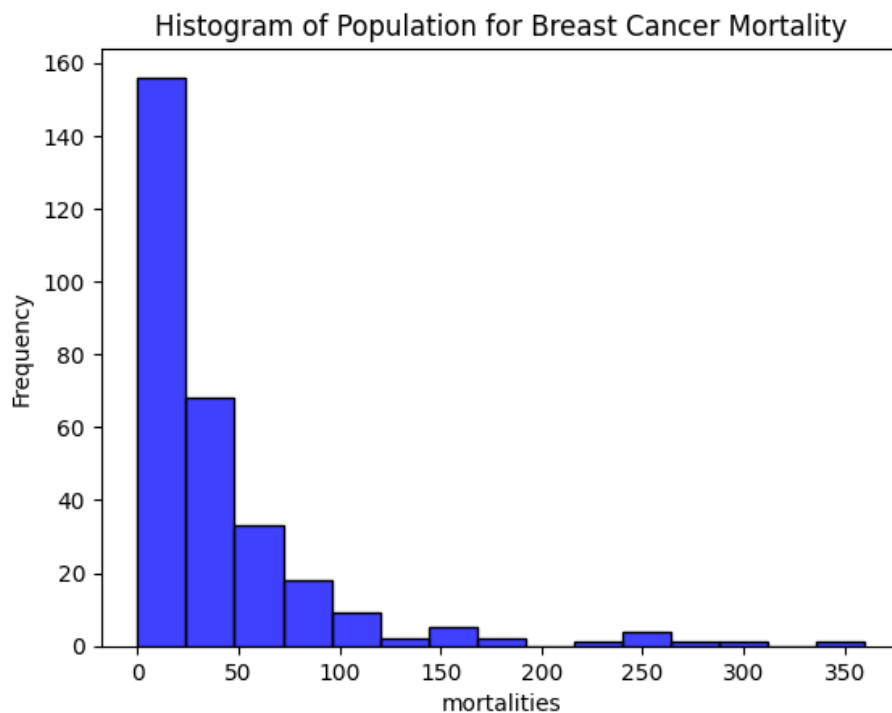
فهرست گزارش سوالات

سوال هفتم: ۳

سوال هشتم: ۷

سوال هفتم:

در اولین قسمت از سوال هفتم به کشیدن نمودار histogram برای داده های در دست می پردازیم. دیده میشود که مناطق محدودی مقدار بسیار زیادی از مرگ و میر دارند ولی اکثر مناطق مرگ و میر محدودی دارند.

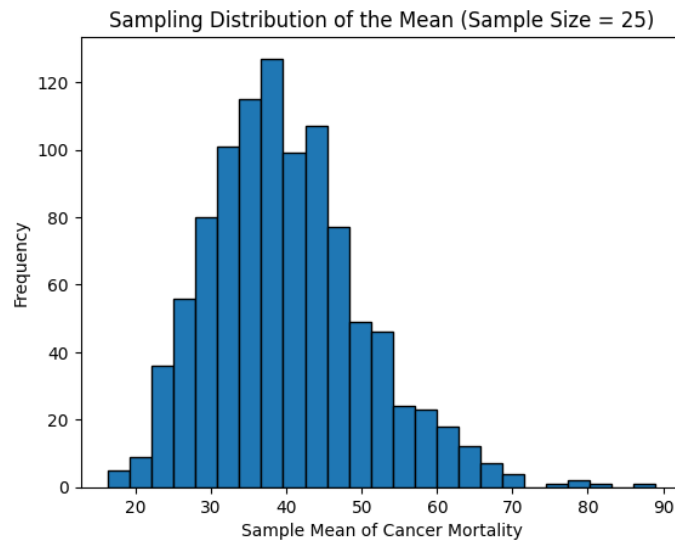


نمودار ۷-۱: توزیع مرگ و میر به دلیل سرطان

در ادامه به بررسی پارامترهای آماری جامعه میپردازیم. با بررسی داده های در دست میتوان دید که میانگین جمعیت در نواحی مختلف عدد ۱۱۲۸۸ است. همچنین کل افراد مرده به دلیل سرطان ۱۱۹۹۷ نفر بودند. برای واریانس جامعه عدد ۱۸۹۲۵۷۸۱۸ و مقدار انحراف معیار عدد ۱۳۷۵۷ خواهد بود.

حال به نمونه برداری از این توزیع می پردازیم و ۱۰۰۰ نمونه ۲۵ تایی از مقادیر مرگ و میر ناشی از سرطان برمیداریم. همانطور که انتظار داشتیم این توزیع مشابه یک توزیع نرمال به ست آمد که مرکز آن حدود ۳۵ است.

بعد از این مرحله کاری که میکنیم این است که یک نمونه ۲۵ تایی از مقادیر مرگ و میر برمیداریم. به کمک این نمونه برداری میانگین تخمین زده شده مقدار ۳۳.۵۲ خواهیم داشت و برای تعداد کل مرگ و میر های بر اساس سرطان ۱۰۰۸۹ خواهیم داشت. به همین ترتیب واریانس نمونه عدد ۱۸۰۱ و انحراف معیار این نمونه ۴۲.۴۴ خواهد بود.



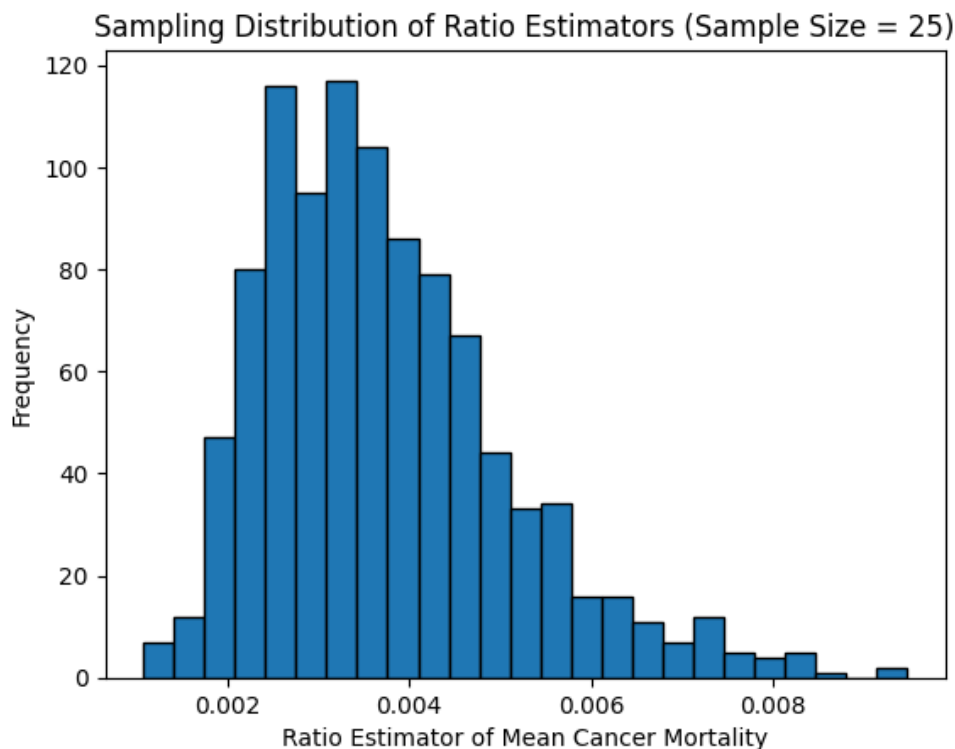
نمودار ۷-۲: توزیع میانگین نمونه‌ها به ازای مقادیر مختلف میانگین

بعد از این مرحله حال به سراغ تخمین بازه ۹۵ درصد برای مقدار میانگین و مقدار کل جمعیت خواهیم رفت. برای بازه اطمینان ۹۵ درصد میانگین نمونه‌های مرگ و میر خواهیم داشت که این بازه برابر (۵۸.۰۸, ۱۸.۶۳) خواهد بود. همچنین برای بازه اطمینان ۹۵ درصد جمعیت کل مرگ و میر خواهیم داشت که برابر (۱۷۴۸۳.۲۷, ۵۶۰۹.۴۴) خواهد بود. میتوان دید که در هر دو مورد تخمین ما شامل مقادیر واقعی یعنی ۳۹.۴۴ و ۱۱۹۹۷ میشود.

پس از موارد ذکر شده نمونه برداری را با ۱۰۰ نمونه ادامه میدهم. مقدار میانگین این نمونه برابر ۳۸.۲۷ و مقدار تخمین زده شده برای کل مرگ و میر ها برابر ۳۸۲۷ است. واریانس نمونه برابر ۲۴۸۸ و همچنین انحراف معیار آن برابر ۴۹.۸۸ میشود. به همین ترتیب بازه های اطمینان ۹۵ درصد برای میانگین (۲۸.۳۷, ۴۸.۱۶) و برای کل مرگ و میر (۱۴۴۹۸, ۸۵۴۰) میشود که باز هم مانند حالت قبل شامل مقادیر واقعی هستند اما بازه کوچکتر شده است.

در مورد اینکه آیا دانستن مقدار جمعیت بخش‌ها باعث افزایش دقت تخمین ما میشود یا خیر میتوان گفت که در صورتی که در بخش‌های با جمعیت بالا توزیع متفاوتی داشته باشیم میتوان گفت که با کمک داشتن جمعیت بخش‌ها تخمین بهتری زد. به عنوان مثال انتظار داریم در مناطق با جمعیت زیاد درصد مرگ و میر ناشی از سرطان بیشتر باشد. به طور کلی داشتن توزیع توأمان دو متغیر در بدترین حالت اطلاعات جدیدی به ما نمیدهد ولی ممکن است دارای اطلاعات مفیدی باشد.

حال به استفاده از تخمین گر های نسبت می پردازیم. در اولین قسمت به بررسی توزیع نسبت ها میپردازیم و با بررسی خروجی ها مشاهده میکنیم که توزیع نسبت ها مشابه توزیع میانگین ها است. همچنین خروجی میانگین تخمین زده شده به کمک روش نسبت هم مشابه خروجی قسمت c است.



نمودار ۸-۲: توزیع نسبت های مرگ به جمعیت

با برداشتن یک نمونه ۲۵ تایی از جمعیت میتوان دید که حدس برای میانگین مرگ و میر به مقدار ۴۱.۵۳ و مقدار جمعیت کل مرگ و میر به ۱۱۸۱۸ تغییر میکند. این مقادیر در قسمت d برابر ۳۳.۵۲ و ۱۰۰۸۹.۵۲ بود. همانطور که دیده میشود به کمک روش نسبت شرایط بهتری برای تخمین داشتیم.

در نهایت برای تشکیل بازه اطمینان خواهیم داشت که بازه ۹۵ درصد برای میانگین مرگ و میر (۳۸.۹۲, ۵۳.۱۷) خواهد بود.

سوال هشتم:

ابتدا کد داده شده را بررسی میکنیم. در کد داده شده دو متغیر تصادفی مجزا داریم و احتمال آن را بررسی خواهیم که مجموع توان دو آنها کمتر از یک باشد. این گزاره معادل آن است که نقطه تصادفی به دست آمده درون یک ربع دایره از یک مربع واحد قرار گرفته باشد که معادل $\frac{\pi^2}{4}$ است. به عبارت دیگر $\frac{\pi^2}{4}$ نسبت نقطه های درون ربع دایره به بیرون خواهد بود. به همین ترتیب با استدلال ذکر شده و همچنین قطعه کد داده شده میتوان دید که مقدار عدد پی ۴ برابر نسبت نقاط درون ربع دایره به کل نقاط خواهد بود.

جدول ۸-۱: مقادیر تخمینی برای عدد پی به ازای تعداد مختلف نقاط

تعداد نقاط	۱۰	۱۰۰۰۰	۱۰۰۰۰۰۰
مقدار تخمین	۳.۲	۳.۱۲۸	۳.۱۴۱۱

همانطور که از نتایج جدول ۸-۱ مشخص است و همچنین انتظار داشتیم هرچه تعداد نقاط برای تخمین را بیشتر کنیم مقدار تخمین به عدد پی نزدیک تر خواهد بود.

در ادامه کار سعی میکنیم با این روش (مونتِه کارلو) مساحت بیضی را حدس بزنیم به این صورت که مستطیلی به اندازه $2a \times 2b$ در نظر میگیریم که در آن a و b شعاع های بزرگ و کوچک بیضی خواهند بود. سپس به تولید نقاط تصادفی میپردازیم و مشاهده میکنیم که چند درصد نقطه ها درون بیضی خواهند بود و سپس این درصد را ضرب در مساحت کل مستطیل میکنیم تا به این صورت مساحت بیضی به دست بیاید.

با تخمین برای ۱۰ میلیون نقطه مساحت بیضی ای با شعاع های ۵ و ۳ را ۴۷.۱۲۵۶ به دست می آوریم که خطای حدود ۰.۰۰۱ دارد.